

TARF: Web Toolkit for Annotating RNA-Related Genomic Features

Authors : Jialin Ma, Jia Meng

Abstract : Genomic features, the genome-based coordinates, are commonly used for the representation of biological features such as genes, RNA transcripts and transcription factor binding sites. For the analysis of RNA-related genomic features, such as RNA modification sites, a common task is to correlate these features with transcript components (5'UTR, CDS, 3'UTR) to explore their distribution characteristics in terms of transcriptomic coordinates, e.g., to examine whether a specific type of biological feature is enriched near transcription start sites. Existing approaches for performing these tasks involve the manipulation of a gene database, conversion from genome-based coordinate to transcript-based coordinate, and visualization methods that are capable of showing RNA transcript components and distribution of the features. These steps are complicated and time consuming, and this is especially true for researchers who are not familiar with relevant tools. To overcome this obstacle, we develop a dedicated web app TARF, which represents web toolkit for annotating RNA-related genomic features. TARF web tool intends to provide a web-based way to easily annotate and visualize RNA-related genomic features. Once a user has uploaded the features with BED format and specified a built-in transcript database or uploaded a customized gene database with GTF format, the tool could fulfill its three main functions. First, it adds annotation on gene and RNA transcript components. For every features provided by the user, the overlapping with RNA transcript components are identified, and the information is combined in one table which is available for copy and download. Summary statistics about ambiguous belongings are also carried out. Second, the tool provides a convenient visualization method of the features on single gene/transcript level. For the selected gene, the tool shows the features with gene model on genome-based view, and also maps the features to transcript-based coordinate and show the distribution against one single spliced RNA transcript. Third, a global transcriptomic view of the genomic features is generated utilizing the Guitar R/Bioconductor package. The distribution of features on RNA transcripts are normalized with respect to RNA transcript landmarks and the enrichment of the features on different RNA transcript components is demonstrated. We tested the newly developed TARF toolkit with 3 different types of genomics features related to chromatin H3K4me3, RNA N6-methyladenosine (m6A) and RNA 5-methylcytosine (m5C), which are obtained from ChIP-Seq, MeRIP-Seq and RNA BS-Seq data, respectively. TARF successfully revealed their respective distribution characteristics, i.e. H3K4me3, m6A and m5C are enriched near transcription starting sites, stop codons and 5'UTRs, respectively. Overall, TARF is a useful web toolkit for annotation and visualization of RNA-related genomic features, and should help simplify the analysis of various RNA-related genomic features, especially those related RNA modifications.

Keywords : RNA-related genomic features, annotation, visualization, web server

Conference Title : ICBBE 2017 : International Conference on Bioinformatics and Biomedical Engineering

Conference Location : Toronto, Canada

Conference Dates : June 15-16, 2017